3966–3976 Nucleic Acids Research, 1994, Vol. 22, No. 19

RNA tertiary structure of the HIV RRE domain II containing non-Watson – Crick base pairs GG and GA: molecular modeling studies

Shu-Yun Le*, N.Pattabiraman[1] and Jacob V.Maizel, Jr
Laboratory of Mathematical Biology, Division of Cancer Biology, Diagnosis and Centers, National Cancer Institute, NIH, Frederick, MD 21702 and [1]Frederick Biomedical Supercomputing Center, Program Resources, Inc., NCI/FCRDC, PO Box B, Frederick, MD 21702, USA

Received May 13, 1994; Revised and Accepted August 8, 1994

ABSTRACT

We have used molecular modeling techniques to model the RNA tertiary structure of the viral RNA element (referred to as domain II of Rev responsive element, RRE) bound by the Rev protein of HIV. In this study, the initial three-dimensional model was built from its established RNA secondary structure, including three non-Watson – Crick G:G, G:A and G:U base pairs. Molecular dynamics (MD) simulations were performed with hydrated or unhydrated sodium ions. Our results indicate that the non-Watson – Crick base pairs in the simulation with unhydrated sodium ions and water are more stable than those with hydrated sodium ions only. The RNA can maintain its compact double helical structure throughout the course of the MD simulations with water and unhydrated sodium ions, although the non-Watson – Crick base pairs and two bulge loops show much more flexibility and conformational distortion than the classical RNA helical region. The distinct distortion of the sugar-phosphate backbone significantly widens the RNA major groove so that the major groove is readily accessible for hydrogen bonding by specific Rev binding. This model emphasizes the importance of specific hydrogen bonding in the stabilization of the three-dimensional structure of the HIV Rev core binding element, not only between the nucleotide bases, but also among the ribose hydroxyls, phosphate anionic oxygens, base oxygens and nitrogens, and bridging water molecules. Moreover, our results suggest that sodium ions play an important role in the formation of base pairs G:G and G:A of the RRE by a manner similar to the arginine of the Rev – RRE complex.

INTRODUCTION

Sequence-specific RNA recognition by proteins is involved in a wide range of regulatory pathways for mRNA processing. For example, gene expression in human immunodeficiency virus (HIV) has been demonstrated to be regulated posttranscriptionally by the RNA binding protein Rev. Without Rev, HIV structural gene transcripts fail to accumulate in the cytoplasm so that the virus cannot replicate (2,3). Although the mechanism of this regulated gene expression is still not clear, considerable progress on the study of the Rev-binding element has been made recently. Malim *et al.* (1) have mapped a *cis*-acting 234 nucleotide (nt) fragment within the *env* gene, referred as the Rev responsive element (RRE), which has been demonstrated to be required for activation by Rev. *In vitro* Rev can bind specifically to the RRE. Of particular significance is the discovery that a 66 nt segment, termed the domain II of RRE (5,9), is necessary and sufficient for high affinity *in vitro* Rev binding (4–7). Also, the domain II alone is sufficient for the Rev response *in vivo* (8). Further studies (10,11) of *in vitro* genetic selection of functional Rev-binding variants from random sequence pools have identified a Rev core binding element (RBE) about 20 nt in length that can bind to Rev as effectively as that of the 234 nt wild-type RRE.

It has been known that the motif for specific recognition or direct binding of an RNA sequence may lie in either the primary, secondary, or tertiary structure of the RNA element. Bartel *et al.* (10) have suggested that non-Watson–Crick base pairs G:A and G:G in a stem-loop (with bulges) structure of the RBE are required for Rev binding *in vitro* and Rev response *in vivo*. Their results have been fully confirmed in another extensive experiment (11) of selective optimization of the Rev-binding element using two independent random sequence pools and validated in the Rev-binding experiment (38) of a series of synthetic RNA duplexes. These results indicate that the distinct distortion of the sugar-phosphate backbone introduced by G:G and G:A may be a critical determinant of recognition by Rev. It seems that local distortion of RNA A-form geometry in RNA double helical region plays an important structural role in the protein binding and recognition.

Fewer studies have been done on the determination of RNA three-dimensional structures. At present, only a few X-ray crystal structures of tRNA molecules (12,13) and RNA double helices (14), as well as nuclear magnetic resonance (NMR) spectra data (15–18) for small stem-loop and hairpin structures have been

*To whom correspondence should be addressed

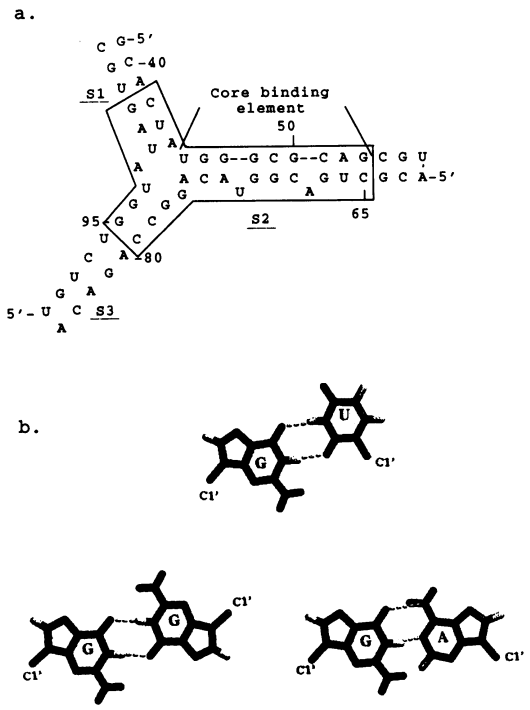


Figure 1. (**a**) RNA secondary structural model of the domain II of HIV RRE, in which two hairpin loops in S1 and S2 are not included. The model is based on the results of Malim *et al.* (1) and Bartel *et al.* (10). The nucleotide numbering in the figure is based on the RNA secondary structure model of HIV RRE (10). The 34 nt RNA fragment in the box is used to model the three-dimensional structure of the HIV RBE. the RBE is delineated by the two upper diagonal lines in the figure. (**b**) Base pairs G:U, G:G and G:A used in the molecular modeling.

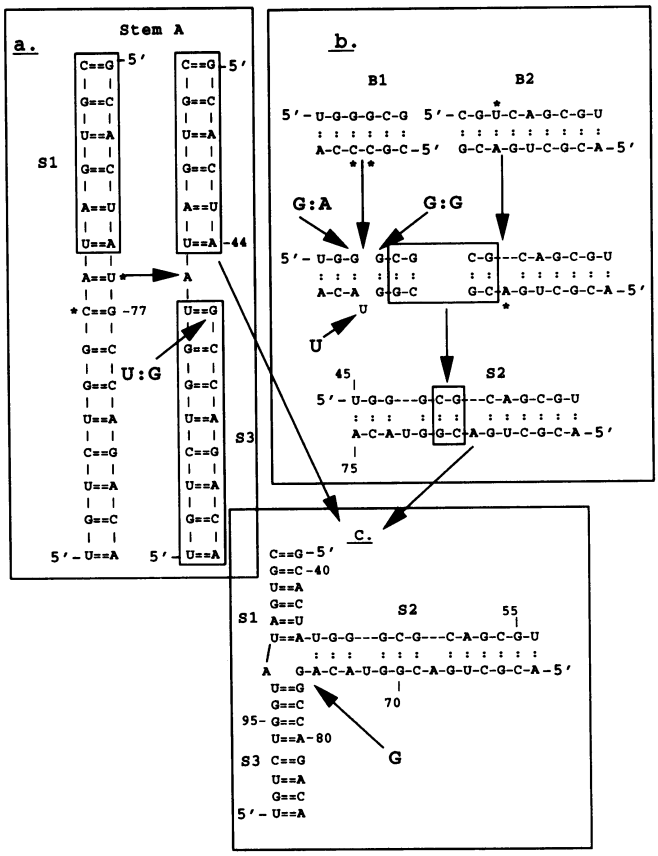


Figure 2. Schematic representation of modeling the initial structure of the domain II of HIV RRE. The modified nucleotide in the modeling is marked by an asterisk. (**a**) Building of stem S1 and stem S3. The marked nt U of the base pair U:A is deleted and the marked base pair G:C is replaced by G:U. (**b**) Building of stem S2. Two marked base pairs G:C in B1 are replaced by G:A and G:G. A bulge U is inserted between G:A and G:G and it is made to be looped outside of B1. The marked nt U of the base pair U:A in B2 is deleted and the base A is made to be stacked inside of B2. The stem B1 including a bulge U is then minimized to release the contacts introduced in the manual modification. The stems B1 and B2 are assembled by superimposing two base pairs which are boxed in the figure. (c) Assembling of stems S1, S2 and S3. In the assembling we open the stacking of S1 and S3, and connecting A44 in S1 and U45 in S2, and connecting S2 and S3 by an additional nt G. (a–c) are boxed, respectively.

determined at high resolution. Recently, the structures of DNA duplex (19) incorporating non-Watson–Crick base pair G:A and RNA double helix including G:U and C:U have been reported (14). The results (14) indicate that the non-Watson–Crick base pairs can be stabilized by water molecules that bridge between ribose hydroxyl and the ring nitrogens. The RNA secondary structure of the HIV RRE has been well established (1,4,9,10,20); the three-dimensional structure, however, is not yet determined on the basis of physical data. It has been recently reported that some aminoglycoside antibiotics, such as neomycin B, can block binding of the Rev protein to RRE (35). These specific inhibitors of the interaction between the Rev protein and RRE recognize the RBE which includes two non-Watson–Crick base pairs G:G and G:A. In order to understand the structural basis of the direct binding between Rev and RBE, and between aminoglycoside antibiotics and RBE (36), it is important to determine the three-dimensional structure of the domain II of HIV RRE. As our first step we modeled the three-dimensional structure of the domain II using known secondary structural constraints, molecular mechanics and MD simulations.

Molecular mechanics and MD simulations are useful theoretical approaches for studying structure and dynamic features of macromolecules. Since the MD method was developed, many theoretical simulations for proteins and DNA double helices have been reported (41); however, only a few studies of simulations for RNA structures (21,40) have been performed. We report here an extensive MD and molecular mechanics simulation for RNA double helices of the domain II of HIV RRE. In the simulations we used fully anionic phosphates with hydrated sodium counterions for the domain II (56 nt; 10 nt in two hairpin loops of the domain II are not enclosed). Based on the structural information derived from the domain II, the further simulations for a 34 nt fragment including the RBE were performed by using 34 sodium ions and 1929 water molecules. From these theoretical simulations we propose the three-dimensional structural model of the RNA fragment in this paper. The structural features of the RBE, suggested by Bartel *et al.* (10), 5'-(45)UGGGCGCA-G(53) and 5'-(65)CUGACGGUACA(75), are analyzed and discussed. Our results suggest that the non-Watson–Crick base pairs G:A and G:G in our model are stabilized by water molecules and sodium ions, which bridge among the oxygens and nitrogens of the purine, phosphate anionic oxygens, and ribose hydroxyl group.

METHODS

In this report, the RNA three-dimensional structures were modeled on a Silicon Graphics workstation by MidasPlus of UCSF (25), a computer graphics system of molecular interactive display and simulation. The molecular mechanics calculation and MD simulation were performed by using AMBER 4.0 (27) on the CRAY Y-MP/8128 supercomputer and the CONVEX 3220 computer system. The force field parameters used in both the molecular mechanics calculation and MD simulation were taken from the standard force field (28) in the AMBER 4.0 package.

Building initial model (M1) of the domain II of HIV RRE

The initial three-dimensional structure of the domain II of HIV RRE was built based on the RNA secondary structure (Figure 1a) proposed by Bartel *et al.* (10). It is a refinement of the original RRE secondary structure (1). Since the two hairpin loops in the domain II are not involved in binding of the HIV Rev protein to its viral RNA recognition element, we did not include them in our modeling studies. The RNA secondary structure was divided into three stems (Figure 1a), termed as S1, 5′-(39)GC-ACUA(44) and 5′-(99)UAGUGC(104), S2, 5′-(45)UGGGCG-CAGCGU(56) and 5′-(62)ACGCUGACGGU-ACA(75) and S3, 5′-(77)GCCAGACA(84) and 5′-(90)UGUCUGGU(97), as well as a multiple loop composed of nt G76 and A98. In the modeling, non-Watson−Crick base pair G:U was generated according to the classical Wobble base pair (24). The base pairs G:A and G:G were generated by their symmetric base-pairing patterns (24). These three non-Watson−Crick base pairs are shown in Figure 1b.

Using the coordinates of the canonical A-form RNA double helical structure derived from the fiber diffraction data (22,23) we first generated a stem A (Figure 2a) that was composed of stem S1, stem S3, and an extra base pair, U:A in the junction of S1 and S2. For the stem A, one should note that the first base pair in the stem S3 is not G:U but G:C. The G:C was then replaced by the G:U by superimposing the base G of G:U on the G of the G:C of stem A. For the extra base pair, U:A, we broke the phosphodiester bond (O3′−P) between the A44 and the U, and phosphodiester bond (P−O5′) between the U and the G77, and deleted the U (Figure 2a). Thus, we built the stacking model of stem S1 and stem S3.

The building of the stem S2 was more complicated than stems S1 and S3 because S2 not only contained two non-Watson−Crick base pairs G:A and G:G but also two bulges with a A and U. Our strategy was to construct two standard A-form RNA double helices B1 and B2 (Figure 2b) at the first step. Then, we replaced two G:C base pairs in the duplex B1 by base pairs G:A and G:G shown in Figure 1b. In the third step, we made a bulge U between G:A and G:G to stack outside of the RNA duplex B1. We deleted the U of the base pair U:A in the stem B2 (marked in Figure 2b) and made the A to be a bulge stacked inside of the RNA duplex B2. The different stacking for bulge U in B1 and bulge A in B2 was made because a pyrimidine base bulge preferred to be stacked outside the helix (30,31) and a purine base bulge preferred to be stacked inside the helix (32−34). At the fourth step, we optimized the model of RNA duplex B1 up to the energy gradient of 0.1 kcal/mol×Å by the program AMBER to release the contacts introduced in the modeling. The energy minimization was performed by confining both the 5′ end base pair U:A and the base pairs C:G and G:C in the 3′ end (i.e. the three base pairs were kept rigid) to ensure proper connections afterwards

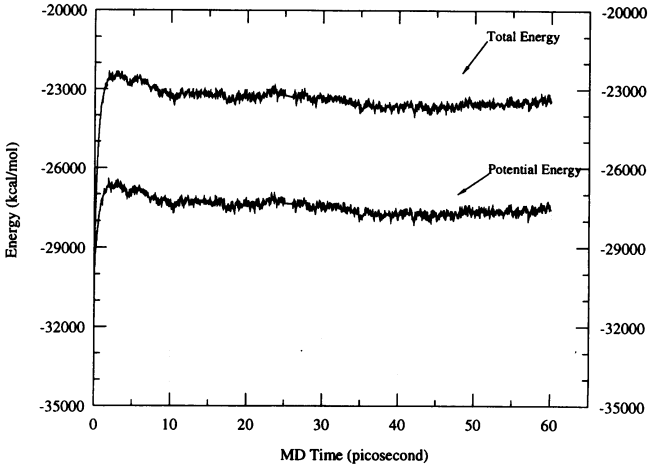


Figure 3. Total energy and potential energy against time for 60 ps of MD simulation in the system including a 34 nt RNA fragment, 34 unhydrated sodium ions and 1929 water molecules.

between stems B1 and B2, as well as between stems S1 and S2, and S2 and S3. In this stage, no sodium ions and water molecules were included in the simulation system. With the minimized structure of RNA duplex B1, we generated the stem S2 by superimposing the base pairs C:G and G:C in the 5′ end of the duplex B2 on the C:G and G:C in the 3′ end of the duplex B1 (Figure 2b). The initial three-dimensional structure model, M1 (model 1) of the domain II of HIV RRE was assembled by connecting the stems S1 and S2, and S2 and S3. The stem S2 was linked to the S1 by the atom O3′ of the nt A44 of S1 and the atom P of the U45 of S2. The S3 was connected to the S2 by an additional G. Stems S1, S2 and S3 were maintained at the same plane. The model M1 was built as compact as possible.

Theoretical simulations for RNA tertiary structures

In the simulations the base pairs in the both 5′ and 3′ ends of the RNA double helices were kept rigid and the 'belly type' of dynamics and energy minimization were used throughout these simulations. In the MD simulations, distance constraints were imposed on all bonds. The simulation temperature was increased slowly from 20 to 298 K with 0.2 ps^{-1} temperature coupling constant. The MD equilibration was carried out up to 8 ps. The system temperature fluctuated around 298 K by ± 10 K during further MD simulations. A time step of 0.001 ps was used during all MD simulations.

Generating model 2 of the RRE domain II structure by the relaxation of M1

The constraints introduced in the initial structure model were expected to release in an extensive MD simulation. Moreover, MD simulations can be used to search a wider range of conformational space for structure refinement. As the limitation of allowable maximum nonbonded pairs in the program MINMD of AMBER 4.0, those simulations for the domain II structure were carried out on the 56 nt RNA fragment with 56 hydrated Na^+ counterions (26). The hydrated Na^+ ions were used to imitate the effect of aqueous hydration of Na^+ and phosphate anion. They were initially placed along the PO_2^- bisector and 6Å from the phosphorus. In the calculation, we used a distance

dependent dielectric function to mimic the presence of a high dielectric solvent, and used a nonbonded cutoff of 20Å. All other parameters were as recommended in the AMBER 4.0 software and documentation. Before starting MD simulations the model, M1, was minimized to the energy gradient of 0.02 kcal/mol×Å.

In the MD simulation we first performed 8 ps of MD equilibration, and then continued the simulation to 100 ps with a 0.2 ps^{-1} temperature coupling constant. After 108 ps MD simulation, the system was further minimized up to the energy gradient of 0.008 kcal/mol×Å. In the energy minimization, those constraints of bonds imposed in the MD simulation were released. This refined structure was termed as the structural model 2 (M2).

Model 3 and further simulation with waters

Based on the structure model M2, we constructed the initial model of the 34 nt RNA fragment by deleting 3 base pairs in the 5′ end of the stem S1, 3 base pairs in the 3′ end of S2, and 5 base pairs in the 3′ end of S3. Thus, the fragment included RBE (10,11) and 2 nt in the multiple loop, 3 base pairs of S1 and 3 base pairs of S3. Further simulations were performed by including 34 Na^+ ions and 1929 water molecules. We put these Na^+ ions on the phosphate bisector 3Å from the P atom, and put the solute RNA into a large water bath that was built by repeating cubes of TIP4P water molecules, which were from a Monte Carlo liquid (35). Those water molecules that were sterically disallowed or were >9Å from any solute atom were removed. No periodic boundary conditions were imposed in the system. The system was first minimized by moving those water molecules and Na^+ ions (that is the RNA fragment was kept rigid) until the energy gradient of the system decreased to 0.01 kcal/mol×Å and then the whole system was further minimized until to 0.1 kcal/mol×Å. In the calculation, we used a constant dielectric function ($\epsilon = 1$) and a nonbonded cutoff of 10Å. The simulation was followed by a 8 ps of MD equilibrated, then a further 52 ps MD simulation at constant temperature of 298 K with a 0.1 ps^{-1} temperature coupling constant. After 60 ps MD simulation, the system was minimized again until the energy gradient decreased to 0.01 kcal/mol×Å. The refined structural model was termed model 3 (M3). During the MD simulation only three water molecules evaporated from the system.

RESULTS AND DISCUSSION

Three-dimensional structures of HIV RBE

Figure 3 is the plot of total energy and potential energy against time in the MD simulation for the model M3. The total energy and potential energy increase rapidly within the heating phase and are relatively stable after equilibration. Throughout the simulation, the water molecules and Na^+ ions were found to change their positions and orientation to stabilize the RNA folding. Significantly, 6 Na^+ ions moved into the grooves of RNA helices and participated in the direct electrostatic interaction with the base keto oxygen and nitrogen, especially for these bases involved in the non-Watson−Crick base pairs. Thirteen Na^+ ions moved close to phosphate oxygens and 15 Na^+ ions moved a short distance away from phosphorus oxygens and formed hydrated Na^+ ions. The Na^+ ions binding to the phosphate oxygens either by the direct interaction or by the bridging water play an important role in the stabilization of the RNA tertiary structure.

In the model M3, the base pairs G47:A73 and G48:G71 were maintained and were quite stable during the MD simulation. All bases in the model M3 are in the anti orientation, even if the *syn* orientation of base G71 in the base pair G48:G71 had been introduced in the starting model, M1. The conformational transition of base G71 from *syn* to *anti* in the MD simulation is shown in Figure 4a. However, the *syn* orientations of bases G71 were observed in the model M2 that was predicted in the simulation system without including water molecules and the unhydrated sodium ions (Figure 4b). Although we also included the hydrated Na^+ in M2, these large hydrated Na^+ ions failed to move into the RNA grooves. They function to neutralize the phosphate anionic oxygens only. This implies that the intermolecular interaction among Na^+ ions, water molecules and nucleosides contributes significantly to the formation of the distinct structural conformation in the stacking region of the two non-Watson−Crick base pairs for the RNA three-dimensional structure. The distributions of glycosidic torsion angle, χ of G48 in MD simulation for M3 and M2 are plotted in Figures 4c and 4d, respectively. The torsion angle, χ is maintained in the *anti* conformation for both M3 and M2 throughout the MD simulation.

The proposed three-dimensional structure (M3) of the 34 nt RNA fragment including the RBE is shown in Figure 5. The backbone torsion angles around the P−$O_{5'}(\alpha)$, $O_{5'}-C_{5'}(\beta)$, $C_{5'}-C_{4'}(\gamma)$, $C_{4'}-C_{3'}(\delta)$, $C_{3'}-O_{3'}(\epsilon)$, and $O_{3'}$−P(ζ), as well as angle χ (around the glycosyl bond) are within the ranges allowed for RNAs (24). Among the backbone torsional angles, angle α and angle γ are more flexible than the others in the simulations and a wide range of conformation changes were observed. The angles β of G47, G48, and A73 are ±ap (antiperiplanar) range as usual (24), the nt G71, however, is −ac (anticlinal) orientation (24) for β. The data of backbone torsional angles of the core binding element in the model M1, M2, and M3 are listed in Table 1a and b.

The conformational transitions of sugar puckering modes in the non-Watson−Crick base pairs were observed. For instance, the conformation of the sugar furanose ring in the nt G47 is $C_2'-exo$ puckering , G48 is $C_2'-endo$, G71 is $C_4'-endo$, and the conformation of A73 is still classic $C_3'-endo$ puckering for the base pairs G47:A73 and G48:G71. In the MD simulation, classical $C_3'-endo$ conformation for RNA double helices are predominant in the Watson−Crick base-pairing region. Significant percentage of $C_1'-endo$ conformation are observed in non-Watson−Crick base-pairing region in the model M3. The RNA double helix of RBE, 5′-(45)UGGGCGCAG(53) and 5′-(65)CUGACGGUACA(75) in M3 is distorted so that it is not in the canonical A-form RNA conformation, with the bending double helix and widening major groove (Figure 5a). The model M3 appears to be a compact and stable RNA three-dimensional structure. The two sugar rings in the bulge loops A68 and U72 show different conformations of $C_2'-exo$ and $C_2'-endo$ puckering, respectively. Both the bulge A68 and bulge U72 are not stacked in the RNA duplex while A68 was initially stacked in the starting structure and the stacking pattern was maintained in the simulation without water and unhydrated sodium ions. The conformation of bulge loop A68 has changed to be looped out in the simulation with water and ions. The two bulge loops cause the helix bending in the RNA duplex of RBE (Figure 5b). The distortion of the sugar-phosphate backbone that was caused by the loops and the base pairs G47:A73 and G48:G71 widened the major groove. These structural features enable the exposed RNA functional groups in the major groove to be accessible for the specific recognition of the Rev protein. Recent results on the RNA recognition by the HIV Tat-derived peptides containing the

arginine-rich region of the Tat protein (37) suggest that the bulge of RNA duplex widens the major groove and permits protein binding and interactive access in the RNA major groove. The structural analogy between RNA-binding protein Rev and Tat and the similar structural roles of their RNA targets suggest that the specific RNA recognition by Rev probably occurs in the major groove of RBE.

Hydrogen bonds and base stacking

The Watson–Crick hydrogen bonds remain nearly intact throughout the simulation in the system with water molecules (Table 2). The direct base-base interactions of G:G and G:A are strengthened under the action of Na^+ and water molecules. In the base pair G47:A73 two hydrogen bonds are between G47(O6) and A73(NH_2) and G47(HN1) and A73(N1). Significantly, a three-center hydrogen bond contributes the direct base-base hydrogen bond of G48:G71 base pairs, in which the donors are HN1 and NH_2 of G48, and the acceptor is O6 of G71. It is interesting that the Lennard-Jones 10–12 potential energies of hydrogen bonds calculated from the base pairs G:G and G:A are −0.93 and −0.76 kcal/mol, respectively. They are as strong as those computed from classic base-pairing in M3. For the model M2, without inclusion of free Na^+ and water molecules, the 10–12 potential energies calculated from the base–base hydrogen bonds of G48:G71, C49:G70 and G50:C69 are −0.07, −0.28, and −0.03 kcal/mol, respectively. This indicates that the hydrogen bonds between these bases are broken and lessened. According to the three-dimensional model of M2, no reasonable base pairing among these bases are observed (data not shown).

The base stacking in the RNA duplex is a main contributor to the thermodynamic stability of the RBE structure. Table 3 shows that the stacking base pairs including a bulge U or bulge

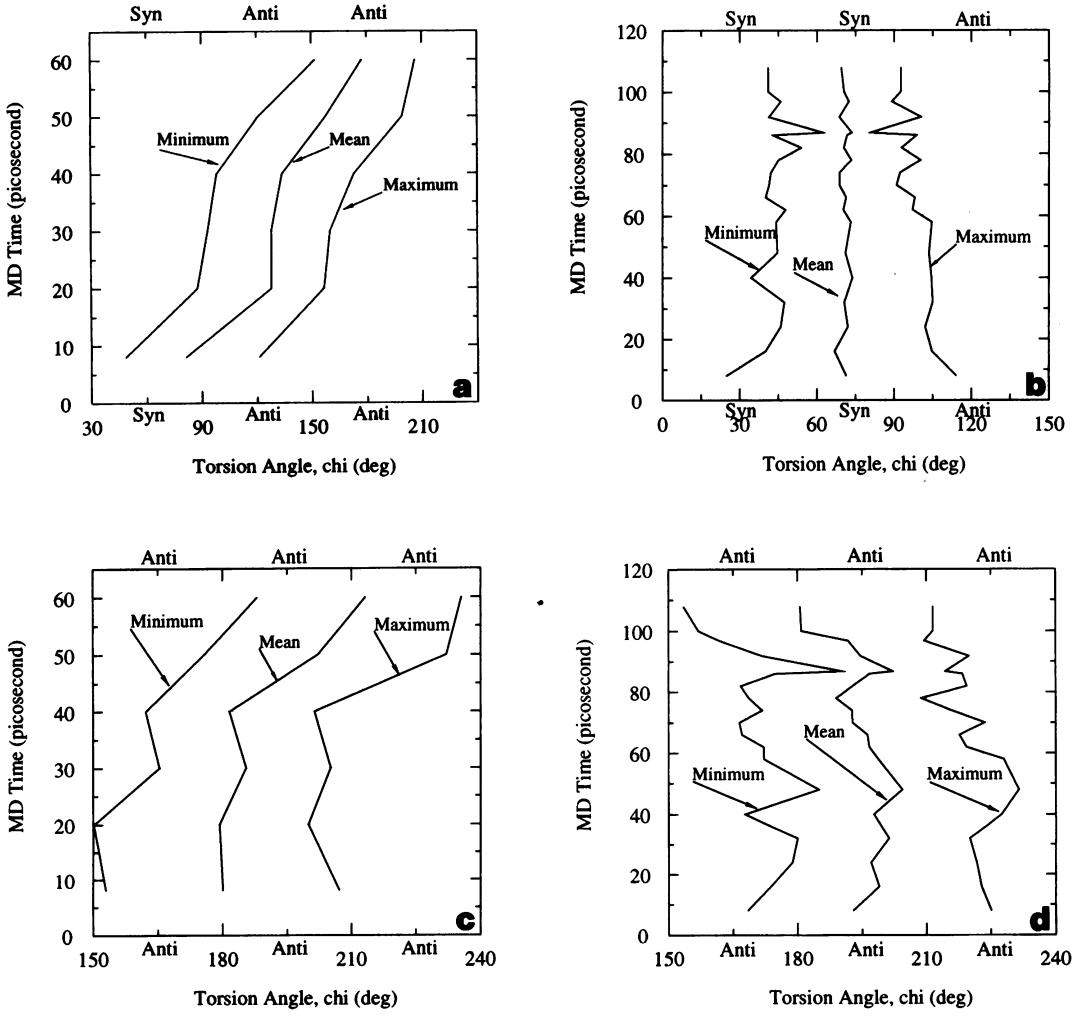
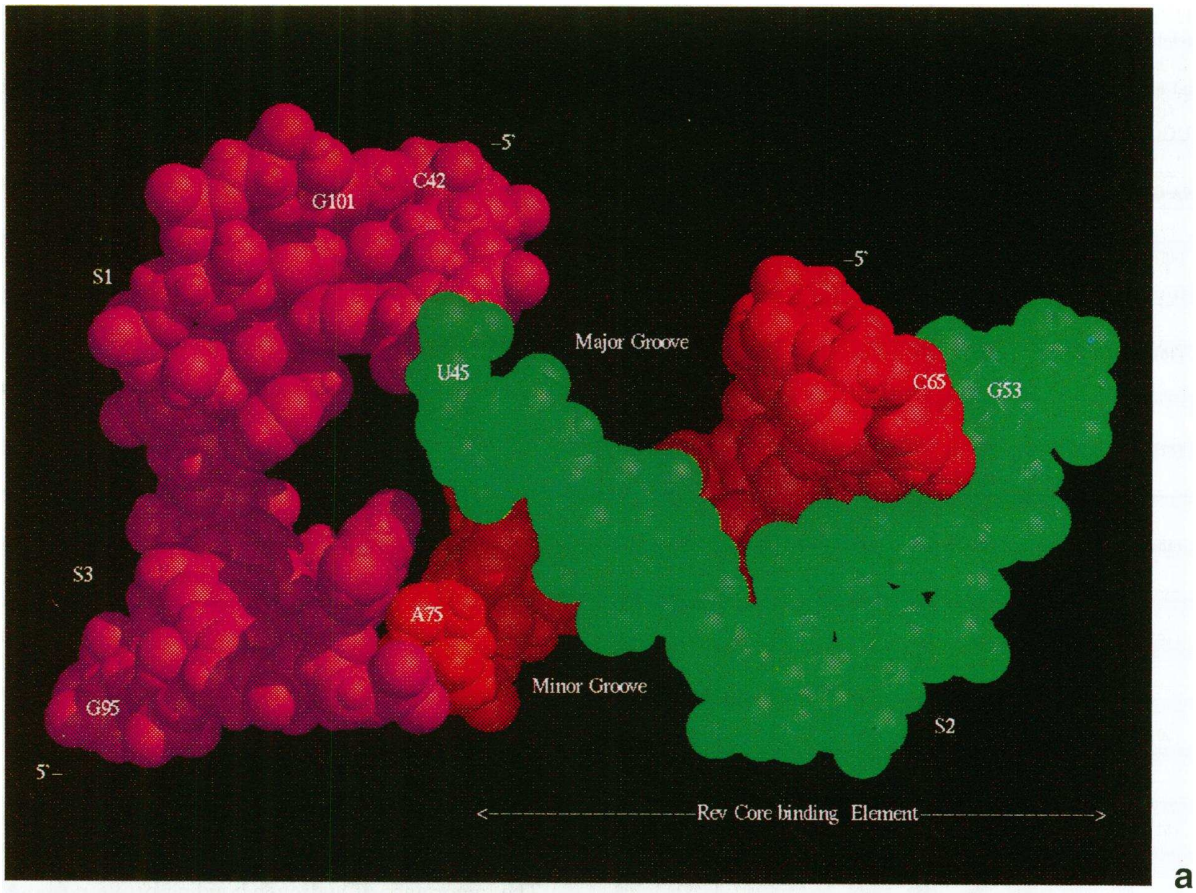
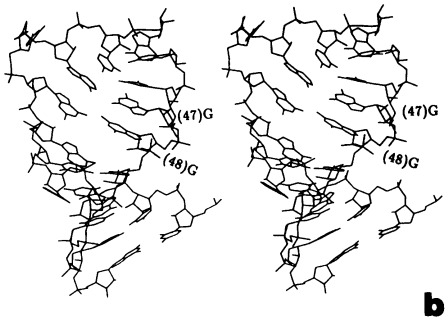


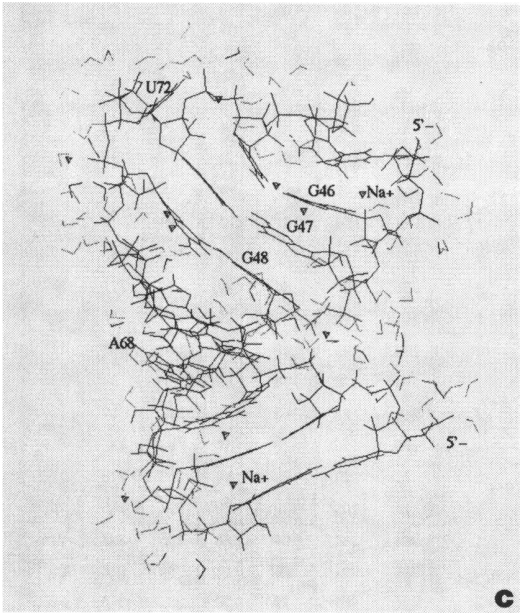
Figure 4. Profiles of the glycosidic torsion angles, χ of G48:G71 in MD simulations for the molecular models M2 and M3. (**a**) Angle χ of G71 in M3 against time during 60 ps MD simulation. In the figure the curves labeled by Minimum, Mean and Maximum indicate the minimum, mean and maximum of angle χ, respectively, calculated from 0 ps to 8 ps, 8 ps to 20 ps, 20 ps to 30 ps, 30 ps to 40 ps, 40 ps to 50 ps and 50 ps to 60 ps. The conformation transition of G71 from *syn* to *anti* was observed after the time of 40 ps (the minimum of angle χ in MD simulation is larger than 90 degree). (**b**) Angle χ of G71 in M2 against time during 108 ps MD simulation. The *syn* conformation was maintained during the MD simulation (see curves labeled by Mean and Minimum). (**c**) Angle χ of G48 in M3 against time during 60 ps MD simulation. The *anti* conformation was maintained during the MD simulation. (**d**) Angle χ of G48 in M2 against time during 108 ps MD simulation. The *anti* conformation was maintained during the MD simulation.



a



b



c

Figure 5. RNA three-dimensional structural model of HIV RRE domain II. (**a**) The CPK model of 34 nt RNA fragment in M3 and without displaying water molecules and sodium ions in which three RNA strains are 5′-(42)CUAUGGGCGCAG(53), 5′-(65)CUGACGGUAC-AGGCC(79) and 5′-(95)GGUAUAG(101). For the RBE, the RNA strain 5′-(45)UGGGCGCAG(53) is shown in green and its complementary strain 5′-(65)CUGACGGUACA(75) is shown in red. The structural model shows a deep and wide major groove in the RBE. (**b**) RNA three-dimensional structure of the RBE in M3 and without displaying water molecules and sodium ions. All base pairs including the base pairs G47:A73 and G48:G71 are stable. (**c**) RNA three-dimensional structure of the RBE in M3 and with water molecules surrounding the phosphate group and sodium ions within 3.0 Å to the RBE. Na^+ is represented by a small triangle. The bulge A68 and U72 are looped out into solution.

Table 1. The RNA backbone torsional angles in the model M1, M2 and M3. M1 (model 1) is the initial structural model of the domain II of HIV RRE, M2 (model 2) is a refined structural model computed including 56 hydrated sodium ions, and M3 (model 3) is the final refined structural model that includes 34 nt, 34 sodium ions and 1927 water molecules. (a) RNA chain of 5′-(45)UGGGCGCA(52); (b) RNA chain of 5′-(66)UGACGGUACA(75)

a

RNA Chain	Model	α	β	γ	δ	ε	ζ	χ
(45)U	1	273	182	60	86	199	295	189
	2	272	181	57	77	171	43	196
	3	262	182	56	83	190	308	198
(46)G	1	278	171	73	83	197	293	191
	2	188	70	163	75	180	282	186
	3	153	170	174	81	169	281	200
(47)G	1	284	165	92	112	283	300	181
	2	174	188	172	116	279	291	187
	3	172	203	175	101	233	275	197
(48)G	1	275	55	171	79	205	296	185
	2	283	69	171	104	211	300	177
	3	148	200	169	152	291	227	214
(49)C	1	268	173	57	80	188	284	217
	2	111	186	173	81	184	292	214
	3	75	269	192	80	188	291	205
(50)G	1	261	179	67	84	204	305	209
	2	147	187	184	77	148	44	183
	3	121	177	181	82	191	297	211
(51)C	1	279	181	65	83	199	289	193
	2	196	172	72	79	182	281	200
	3	153	186	176	91	202	297	188
(52)A	1	276	182	64	83	202	293	206
	2	168	189	167	98	251	284	188
	3	165	165	170	108	236	288	189

b

RNA Chain	Model	α	β	γ	δ	ε	ζ	χ
(66)U	1	267	181	69	78	191	291	205
	2	159	185	167	87	219	287	196
	3	158	194	181	102	210	278	198
(67)G	1	282	174	70	81	186	280	203
	2	281	166	66	80	174	271	192
	3	247	183	67	79	211	301	210
(68)A	1	275	184	69	84	196	293	214
	2	285	178	73	69	156	34	183
	3	249	65	168	89	170	50	118
(69)C	1	275	170	70	87	190	295	216
	2	77	197	180	80	183	270	196
	3	76	202	179	78	191	285	207
(70)G	1	254	187	68	88	285	328	210
	2	281	201	68	111	290	286	187
	3	276	169	72	112	261	310	225
(71)G	1	304	340	181	60	59	179	85
	2	89	204	176	83	39	156	66
	3	198	220	48	155	239	167	192
(72)U	1	129	87	317	85	165	151	210
	2	278	252	63	75	162	154	202
	3	273	79	161	160	282	114	203
(73)A	1	218	153	70	85	190	274	195
	2	69	178	190	78	186	282	192
	3	35	155	197	78	179	279	202
(74)C	1	281	189	74	142	222	303	210
	2	178	197	164	155	278	298	214
	3	168	214	172	141	238	291	201
(75)A	1	197	174	76	70	58	110	238
	2	282	131	54	88	47	150	272
	3	222	182	60	83	21	192	254

In the table, α: $(n-1)O_{3'}-P-O_{5'}-C_{5'}$, β: $P-O_{5'}-C_{5'}-C_{4'}$, γ: $O_{5'}-C_{5'}-C_{4'}-C_{3'}$, δ: $C_{5'}-C_{4'}-C_{3'}-O_{3'}$, ε: $C_{4'}-C_{3'}-O_{3'}-P$, ζ: $C_{3'}-O_{3'}-P-O_{5'}(n+1)$, and χ: $O_{1'}-C_{1'}-N_9-C_4$ for purines A and G, and $O_{1'}-C_{1'}-N_1-C_2$ for pyrimidines C and U.

A are two of the strongest base stacking regions in the structure. It indicates that an insertion of the bulge U looped out into solution between two stacking non-Watson–Crick base pairs G47:A73 and G48:G71 strengthens the stability of the base-pairing stacking. Based on the analysis of stacking energy, the stem S3 is stacked with either S1 and S2. The stacking patterns between stems S2 and S3 (−11.27 kcal/mol) and S1 and S3 (−9.57 kcal/mol) are more stable than that between stems S2 and S1 (0.02 kcal/mol).

Effects of water and Na^+ in the stabilization of the RBE structure

The importance of hydrogen bonding between ribose hydroxyls, phosphate oxygens and water in stabilization of the intramolecular and intermolecular structures of double helical RNA has been recently demonstrated in the crystal structure of RNA duplex $(r-GGACUUCGGUCC)_2$ (14). To focus on the water effects in stabilization of the RBE structures, we have analyzed in more detail the structure of the RNA$-Na^+-H_2O$ interactions. Many intermolecular interactions through water mediated hydrogen bonding are observed. Most of the direct hydrogen bonds involve the phosphate anionic oxygen as an acceptor. Significantly, a ribose O_2' hydroxyl can serve either as a donor or an acceptor in the hydrogen bond with water molecules. This feature observed in our simulations is consistent with that observed in the X-ray structure of RNA double helix containing the non-Watson–Crick base pairs G:U and C:U (14).

The bridging water molecules between phosphate anionic oxygens and Na^+, between oxygen and nitrogen atoms of base rings occurred frequently in the predicted model, M3. Also, the three-center hydrogen bonds by bridging water molecules frequently occurred in this RNA three-dimensional structure. A typical example of these multiple center hydrogen bonds occurred in the base stacking region of the non-Watson–Crick base pair G47:A73, G48:G71, and bulge U72.

Effects of water molecules and Na^+ *in the G:A base pair.* Figure 6a shows the structure of the G47:A73 pair including Na^+, and water molecules bound in the grooves. Two base–base hydrogen bonds form in the base pair G:A. One bridging water molecule between the endocyclic nitrogen N3 and ribose hydroxyl, O_2' of both the guanine and adenine, in the minor groove, seems to provide the stabilization of their specific base orientation. The

Table 2. Hydrogen bonds formed between base pairs in the HIV RBE. The unit of distance is the angstrom, angle is degree. and energy is kcal/mol. The Lennard-Jones 10−12 potential energy of hydrogen bonding is listed in the fifth column. The total interaction energy of base pairing is listed in the sixth column

Base-pairings	Hydrogen Bond Atom Type	Model 3 Dist	Angle	12-10 Energy	Total Energy
(45)U--A(75)	(45)O4--HN6(75)	2.57	140.7	-0.12	-10.10
	HN3--N1	1.85	162.3	-0.18	
(46)G--C(74)	(46)O6--HN4(74)	1.89	138.2	-0.36	-20.09
	HN1--N3	1.87	165.4	-0.27	
	HN2--O2	1.91	164.2	-0.40	
(47)G--A(73)	(47)O6--HN6(73)	1.88	160.8	-0.30	-12.61
	HN1--N1	1.94	165.0	-0.46	
	HN2--N1	-	-	-	
(48)G--G(71)	48)HN1--O6(71)	1.90	149.0	-0.48	-11.88
	HN2--O6	1.93	146.7	-0.45	
(49)C--G(70)	(49)HN4--O6(70)	1.96	148.9	-0.49	-18.82
	N3--HN1	1.94	150.2	-0.46	
	O2--HN2	1.94	138.3	-0.46	
	O2--HN1	2.24	129.8	-0.28	
(50)G--C(69)	(50)O6--HN4(69)	2.06	163.9	-0.48	-19.67
	HN1--N3	1.89	163.9	-0.34	
	HN2--O2	1.86	150.4	-0.22	
	HN1--O2	3.01	122.8	-0.03	
(51)C--G(67)	(50)N3--HN1(67)	1.92	161.2	-0.44	-20.88
	N3--HN2	-	-	-	
	O2--HN2	1.88	154.7	-0.33	
	HN4--O6	1.89	176.9	-0.35	
(52)A--U(66)	(52)HN6--O4(66)	1.98	170.9	-0.50	-10.91
	N1--HN3	1.85	158.5	-0.19	

bridging water in the adenine is also a donor and bond with the atom O_1' of C74 (H-bond is not shown in the figure). In addition to the two three-center hydrogen bonds, another three-center hydrogen bond is composed of the guanine $-NH_2$ group and two water molecules, which are bridged by a Na^+. The Na^+ seems to contribute to the stabilization of this unusual base pairing rather than to neutralize anionic phosphates. This is an important structural feature for the three non-Watson−Crick base pairs G:U, G:A and G:G, because five of the total six Na^+ ions binding to the base in M3 are involved in the direct binding to the bases of G46:C74, G47:A73, G48:G71 and G77:U97. It indicates the importance of those sodium ions in the stabilization of these specific conformations. The water molecules surrounding the G47 and A73 in the minor grooves are bridged together. It is clear that the thermodynamic stability of the non-Watson−Crick base pair G:A is strengthened by these bridging water molecules and the Na^+ among the G47(O_2'), G47(N3), G47(NH_2), A73(N3), and A73(O_2'). The three-center hydrogen bond in the major groove consists of a water molecule and $-NH_2$ group and N7 of the adenine, in which the water molecule serves both as a donor and an acceptor. The structure is stabilized by another bridging water molecule that is involved in the hydrogen bond with $-NH_2$ group of C74 (H-bond is not shown in the figure).

Four Na^+ ions surrounding the two phosphate groups form the hydrated counterions rather than interact directly with the oxygens of the phosphate group. In the model M3, 15 Na^+ ions of 34 Na^+ ions are such hydrated counterions that are bridged by water molecules. Thirteen Na^+ ions can interact directly with the phosphate oxygens.

Table 3. Stacking energy between base-pairings in the model M3. The interacting energy does not include the contribution from water molecules and sodium ions

Base Stacking	Energy (kcal/mol)
(42)C--G(101) (43)U--A(100)	-17.74
(43)U--A(100) (44)A--U(99)	-17.80
(44)A--U(99) A(98)	-9.55
(44)A--U(99) A(98) (77)G--U(97)	-9.57
(44)A--U(99) (45)U--A(75)	0.02
(45)U--A(75) G(76)	-11.24
(45)U--A(75) G(76) (97)U--G(77)	-11.27
(98)A G(76) (97)U--G(77)	-0.04
(45)U--A(75) (46)G--C(74)	-13.10
(46)G--C(74) (47)G--A(73)	-18.21
(47)G--A(73) U(72) (48)G--G(71)	-19.46
(48)G--G(71) (49)C--G(70)	-6.71
(49)C--G(70) (50)G--C(69)	-19.37
(50)G--C(69) A(68) (51)C--G(67)	-19.43
(51)C--G(67) (52)A--U(66)	-16.69
(52)A--U(66) (53)G--C(65)	-17.37
(77)G--U(97) (78)C--G(96)	-17.64
(78)C--G(96) (79)C--G(95)	-15.38

Effects of water molecules and Na^+ *in the G:G base pair.* The structure of the G:G base pair in the model M3 shows an unusual structural feature (Figure 6b). To incorporate a G:G base pair into the RNA helix without introducing severe radial distortion, the relative movement between two bases form a three-center base−base hydrogen bond among the donors G48(HN1) and G48(NH_2), and acceptor G71(O6) even though two are possible. The unusual three-center base−base hydrogen bond of G:G can widen the RNA major groove. The stability of the G:G base pair is further strengthened by including two bridging water molecules between the G48(O6) and G71(HN1) and by including two bridging water molecules and a Na^+ between the G48(NH_2) and G71(N7). The orientation of base G71 has been changed from its initial *syn* to *anti* orientation. In the conformation transition the adjacent bulge U72 plays a critical role in the conformational

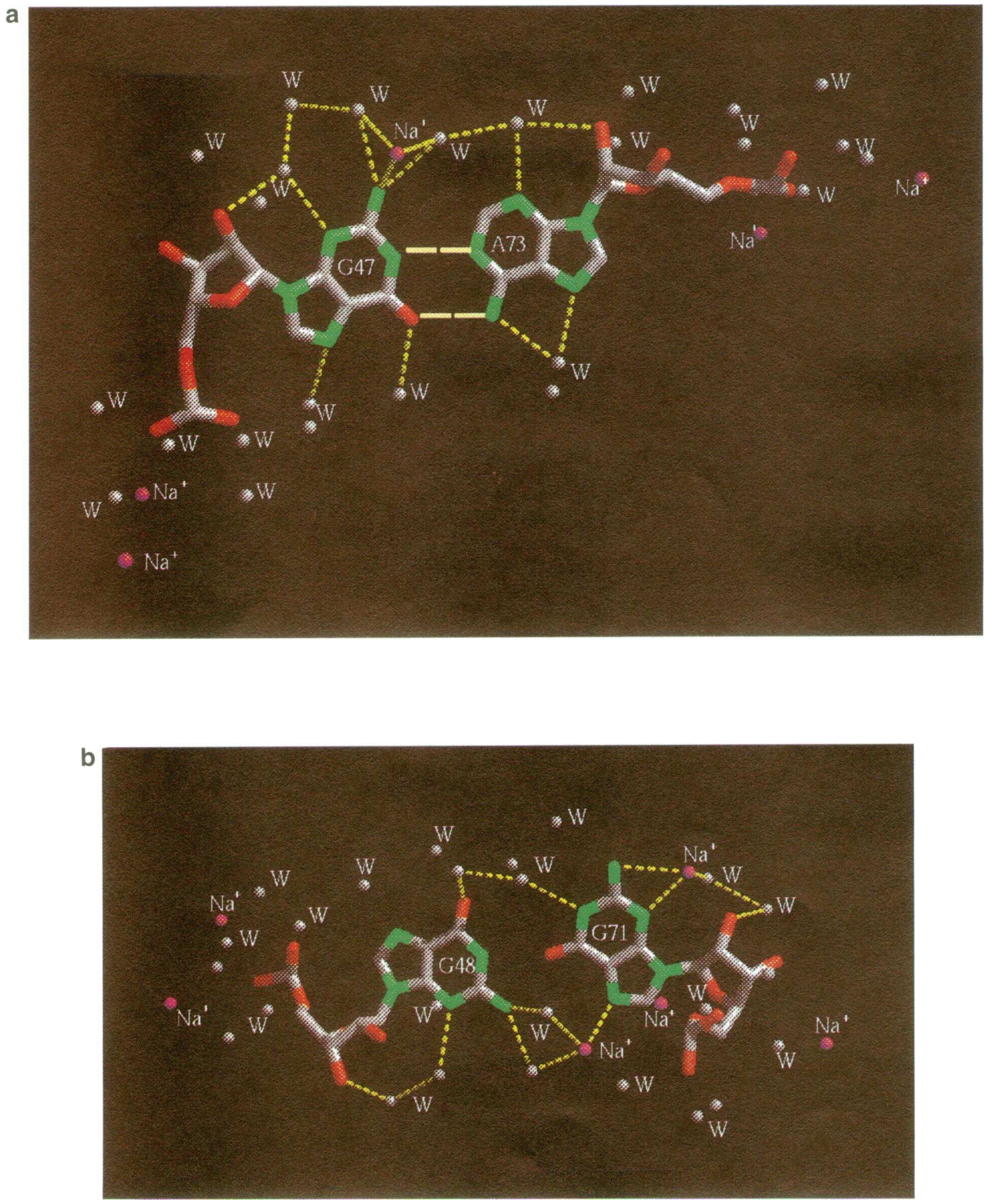


Figure 6. The structures of non-Watson–Crick base pairs G47:A73 in (**a**), and G48:G71 in (**b**), with including bound waters in the predicted model M3. Nitrogens and oxygens in the nt G and A are shaded by green and red, and hydrogens in the G and A are not shown in the figures. Na^+ is shown in magenta ball and labeled. The water molecule is represented by its oxygen is indicated by a white ball (labeled by W). Water–water and water–nucleotide hydrogen bonds are shown as short dashed lines. The direct electrostatic interactions between the Na^+ and nucleotide and between the Na^+ and water molecule are also indicated by short dashed lines. Two base–base hydrogen bonds are indicated by two long dashed lines. Five water–sugar, seven water–water and 13 water–phosphate oxygen hydrogen bonds are not shown in (a), because of the limitation of the program MIDAS (16 distances between specified atoms can be displayed at most in the modeling session). Similarly, the three-center hydrogen bond of G48:G71, three base–water, four sugar–water, one water–water and 12 water–phosphate oxygen hydrogen bonds are also not shown in (b).

transition of base G71, providing more freedom from the constraints of the polynucleotide backbone rotation in the stacking region of the two non-Watson−Crick base pairs. Moreover, the two bridging water molecules and a Na^+ ion among G71(NH_2), G71(N3), and G71(O_2') can play an important role in promoting the conformation transition and in stabilizing the *anti* orientation of the G71. Similarly, the other two bridging water molecules between the G48(N3) and G48(O_3') also contribute to the stabilization of the anti orientation of the G48.

However, the stable base pair G:G is not observed in M2 that was generated with the inclusion of hydrated Na^+ ions but without water. In the simulation of M2, the hydrated Na^+ ions had local motion. These hydrated Na^+ ions surrounded the anionic phosphates and their average distance from the phosphate anionic oxygen was 6.21Å. No hydrated Na^+ ions are observed inside the RNA grooves. They only play a role in the neutralization of the anionic phosphates of the RNA molecule. Taken together our results suggest that those Na^+ ions and bridging water molecules inside the RNA grooves in the model M3 play a crucial role in the stabilization of non-Watson−Crick base pairs in the RBE. These sodium ions can contribute to the stabilization of the distinct structure by bridging oxygens and nitrogens of the base ring and water molecules, and/or by electrostatic interaction directly with these oxygens and nitrogens besides neutralizing anionic phosphates.

Although our initial model was built based on the well established RNA secondary structure of RRE (1,10) uncertainty of RNA tertiary interaction was also introduced in the modeling since we have little information available about its tertiary structure. For example, we do not know what the correct stacking pattern among stems S1, S2 and S3 is. The biased stacking of S1 and S3 over S1 and S2 and over S2 and S3 was imposed in the model building. The molecular simulation results, however, show that the stacking of S2 and S3 is the most stable among these three stacking patterns. It is possible that the stacking of S2 and S3 is a tertiary structural feature of HIV RRE domain II structure. In our studies we are interested in the structural features of two non-Watson−Crick base pairs and RBE that are located at the stem S2. Their structural features are mainly dependent on the tertiary structure of S2 and are less dependent on the stem stacking among S1, S2 and S3. Thus, our results about the RBE structure and its structural features can provide useful information for further studies on the RBE structure and the interaction among the HIV Rev protein, aminoglycoside antibiotics and their viral RNA recognition element.

While preparing this manuscript, a recent publication (39) was brought to our attention. The NMR data of a specific 1:1 complex between RRE stem IIB and the modified Rev_{34-50} peptide indicate that the different structural features occurred in the free RNA and bound RNA. The conformational change in the RRE from free RNA to bound RNA by Rev protein involves unstacking of the bulge loop A68 and the formation of the four additional base pairs, G47:A73, G48:G71, C49:G70 and G50:C69, in the internal loop. Moreover, in the G:G base pair, neither G48 or G71 has *syn* conformation where one of the torsion angles (χ) was previously supposed to be in a *syn* conformation in order to form the proposed G:G base pair (11,38). Our results are compatible with these NMR data. This structural feature in the RBE structure has also been observed in a recent molecular modeling (43) by using different modeling strategy. However, the late NMR studies (41) of a 30-nt RNA fragment, which contains the HIV RBE indicated that G48 and G71 form an *anti/syn* base pair. To improve our model the further molecular simulations using these NMR data (39,42) are necessary.

The proposed structural model of the RNA fragment including 34 nt of RRE domain II emphasizes the importance of these specific hydrogen bonds in the stabilization of the RNA three-dimensional structure. The base pairs, especially, for G:G and G:A are stabilized by bridging water molecules and sodium ions which bridge between bases and between the base and sugar ring. The bending RNA helix and the conformational distortion of sugar-phosphate backbone in the base pairs G:G and G:A widen the RNA major groove of RBE so as to be more accessible for specific binding of the Rev protein. We suggest that these sodium ions binding to the non-Watson−Crick base pairs in our model play a crucial role in promoting the conformational change in the RBE. They act in a manner analogous to the basic amino acids within the arginine-rich domain of Rev in the Rev−RRE complex, which result in the formation of additional base pairs including G:G and G:A base pairs not present in the free RNA.

ACKNOWLEDGEMENTS

We thank Dr Hugo M.Martinez for his interesting discussion on RNA folding and modeling, and his critical comments on this manuscript, and Dr Sergei Gulnik for bringing the paper of Battiste *et al.* (39) to our attention. We acknowledge the Frederick Biomedical Supercomputing Center for allocation of computing time and staff support. Also, a special thanks to Wayne A.Main of FBSC for proofing this manuscript. The content of this publication does not necessarily reflect the views or policies of the Department of Health and Human Services, nor does mention of trade names, commercial products, or organizations imply endorsement by the US Government.

REFERENCES

1. Malim, M.H., Hauber, J., Le, S-Y., Maizel, J.V., and Cullen, B.R. (1989) *Nature* **338**, 254−257.
2. Sodroski, J., Goh, W.C., Rosen, C., Dayton, A., Terwillinger, E., and Haseltine, W.A. (1986) *Nature* **321**, 412−417.
3. Feinberg, M.B., Jarrett, R.F., Aldovini, A., Gallo, R.C., and Wong-Staal, F. (1986) *Cell* **46**, 807−817.
4. jems, J., Calnan, B.J., Frankel, A.D., and Sharp, P.A. (1992) *EMBO J.* **11**, 1119−1129.
5. Malim, M.H., Tiley, L.S., McCarn, D.F., Rusche, J.R., Hauber, J., and Cullen, B.R. (1990) *Cell* **60**, 675−683.
6. Heaphy, S., Dingwall, C., Ernberg, I., Gait, M.J., Green, S.M., Karn, J., Lowe, A.D., Singh, M., and Skinner, M.A. (1990) *Cell* **60**, 685−693.
7. Holland, S.M., Ahmad, N., Maitra, R.K., Wingfield, P., and Venkatesan, S. (1990) *J. Virol.* **64**, 5966−5975.
8. Huang, X., Hope, T.J., Bond, B.L., McDonald, D., Grahl, K., and Parslow, T.G. (1991) *J. Virol.* **65**, 2131−2134.
9. Le, S-Y., Malim, M.H., Cullen, B.R., and Maizel, J.V. (1990) *Nucleic Acids Res.* **18**, 613−1623.
10. Bartel, D.P., Zapp, M.L., Green, M.R., and Szostak, J.W. (1991) *Cell* **67**, 529−536.
11. Giver, L., Bartel, D.P., Zapp, M.L., Pawul, A., Green, M.R., and Ellington, A.D. (1993) *Nucleic Acids Res.* **21**, 5509−5516.
12. Kim, S-H., Suddath, F.L., Quigley, F.L., Mvpherson, A., Sussman, J.L., Wang, A.H.J., Seeman, N.C., and Rich, A. (1974) *Science* **185**, 435−440.
13. Robertus, J.D., Ladner, J.E., Finch, J.T., Rhodes, D., Brown, R.S., Clark, B.F.C., and Klug, L. (1974) *Nature* **250**, 546−551.
14. Holbrook, S.R., Cheong, C., Tinoco, I., and Kim, S-H. (1991) *Nature* **353**, 579−581.
15. White, S.A., Nilges, M., Huang, A., Brunger, A.T., and Moore, P.B. (1992) *Biochemistry* **31**, 1610−1620.
16. Puglisi, J.D., Tan, R., Calnan, B.J., Frankel, A.D., and Williamson, J.R. (1992) *Science* **257**, 76−80.

17. Varani, G., Cheong, C., and Tinoco, I. (1991) *Biochemistry* **30**, 3280–3289.
18. Heus, H.A. and Pardi, A. (1991) *Science* **253**, 191–194.
19. Nikonowicz, E.P. and Gorenstein, D.G. (1990) *Biochemistry* **29**, 8845–8858.
20. Dayton, E.T., Powell, D.M., and Dayton, A.I. (1989) *Science* **246**, 1625–1629.
21. Yao, S. and Wilson, W.D. (1992) *J. Biomol. Struct. Dynam.* **10**, 367–387.
22. Arnott, S., Hukins, D.W.L., Dover, S.D. Fuller, W., and Hodgson, A.R. (1973) *J. Mol. Biol.* **81**, 107–122.
23. Chandrasekaran, R. and Arnott, S. (1989) In: *Landolt-Bornstein, Nucleic Acids*, VII/1b (Saenger, W., ed.). Springer-Verlag, Berlin, 55.
24. Saenger, W. (1984) *Principles of Nucleic Acid Structure.* Springer-Verlag, New York.
25. Ferrin, T.E., Huang, C.C., Jarvis, L.E., and Langridge, R. (1988) *J. Mol. Graphics* **6**, 13–27; MidasPlus User's Manual, Computer Graphics Laboratory, School of Pharmacy, University of California, San Francisco (1992).
26. Singh, U.C., Weiner, S.J., and Kollman, P. (1985) *Proc. Natl Acad. Sci. USA* **82**, 755–759.
27. Pearlman, D.A., Case, D.A., Caldwell, J.C., Seibel, G.L., Singh, U.C., Weiner, P., and Kollman, P.A. (1991) AMBER 4.0, University of California, San Francisco.
28. Weiner, S., Kollman, P.A., Case, D.A., Singh, U.C., Ghio, C., Alagona, G., Profeta, S., and Weiner, P. (1984) *J. Am. Chem. Soc.* **106**, 765–784.
29. Lee, W.K., Gao, Y., and Prohofsky, E.W. (1984) *Biopolymers* **23**, 257–270.
30. Kalnik, M.W., Norman, D.G., Zagorski, M.G., Swann, P.F., and Patel, D.J. (1989) *Biochemistry* **28**, 294–303.
31. Morden, K.M., Gunn, B.M., and Maskos, K. (1990) *Biochemistry* **29**, 8835.
32. Nikonowicz, E.P., Meadows, R.P., and Gorenstein, D.G. (1990) *Biochemistry* **29**, 4193–4204.
33. Rosen, M.A., Live, D., and Patel, D.J. (1992) *Biochemistry* **31**, 4004–4014.
34. Rosen, M.A., Shapiro, L., and Patel, D.J. (1992) *Biochemistry* **31**, 4015–4026.
35. Jorgensen, W., Chandrasekhar, J., Madura, J., Impey, M., and Klein, R. (1983) *J. Chem. Phys.* **79**, 926–935.
36. Zapp, M.L., Stern, S., and Green, M.R. (1993) *Cell* **74**, 969–978.
37. Weeks, K.M. and Crothers, D.M. (1991) *Cell* **66**, 577–588.
38. Iwai, S., Pritchard, C., Mann, D.A., Karn, J., and Gait, M. (1992) *Nucleic Acids Res.* **20**, 6465–6472.
39. Battiste, J. L., Tan, R., Frankel, A.D., and Williamson, J.R. (1994) *Biochemistry* **33**, 2741–2747.
40. Prabhakaran, M., Harvey, S.C., Mao, B. and McCammon, J.A. (1983) *J. Biomol Struct. Dynam.* **1**, 357.
41. McCammon, J.A. and Harvey, S.C. (1987) *Dynamics of Protein and Nucleic Acids.* Cambridge University Press, Cambridge.
42. Peterson, R.D., Bartel, D.P., Szostak, J.W., Horvath, S.J., and Feigon, J. (1994) *Biochemistry* **33**, 5357–5366.
43. Leclerc, F., Cedergren, R., and Ellington, A.D. (1994) *Nature Struct. Biol.* **1**, 293–300.